NOVETTA

Impact of Non-Facial Regions on Face Recognition Performance

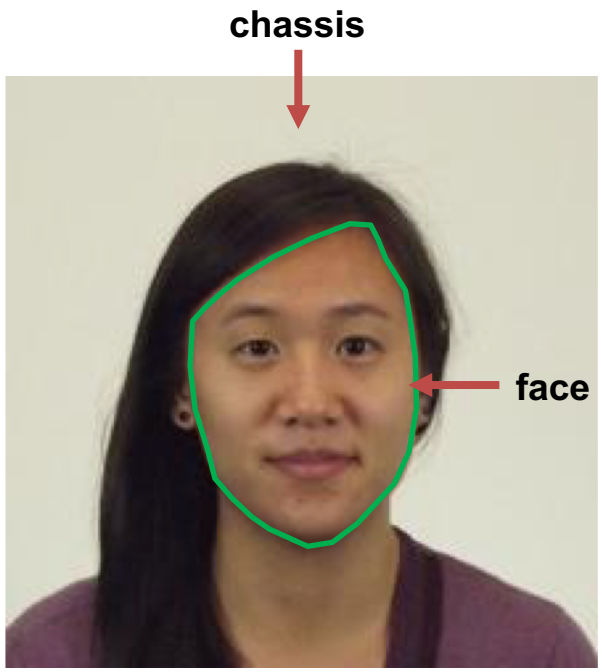
Michelle Jou
Kyle O'Brien
Michelle Rybak
Michael Thieme

2018 INTERNATIONAL FACE PERFORMANCE CONFERENCE

29 November 2018

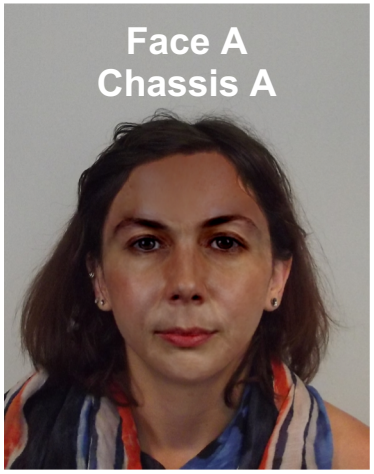
Background

- Novetta conducts biometric testing and builds automated face synthesis software
- Our work is heavily reliant on precise segmentation of the face region
- We frequently observe anomalous behaviors in face recognition (FR) systems
  - Anomalies: instances where FR behavior does not align with visual expectations
- This briefing examines whether FR tools use non-face regions ("chassis") for recognition
  - A subject's chassis can vary across photos (hair styles, clothing, etc.), such that it is not suitable for determining identity

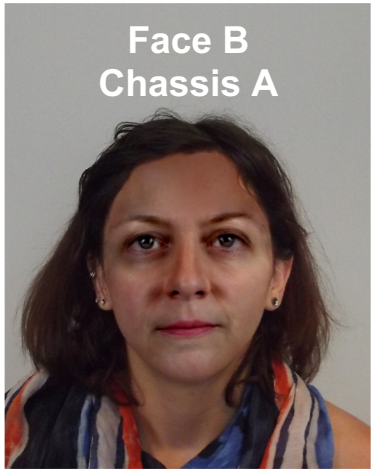


Hypothesis

- In the synthesis context, two images that share the same chassis but different faces should generate low comparison scores

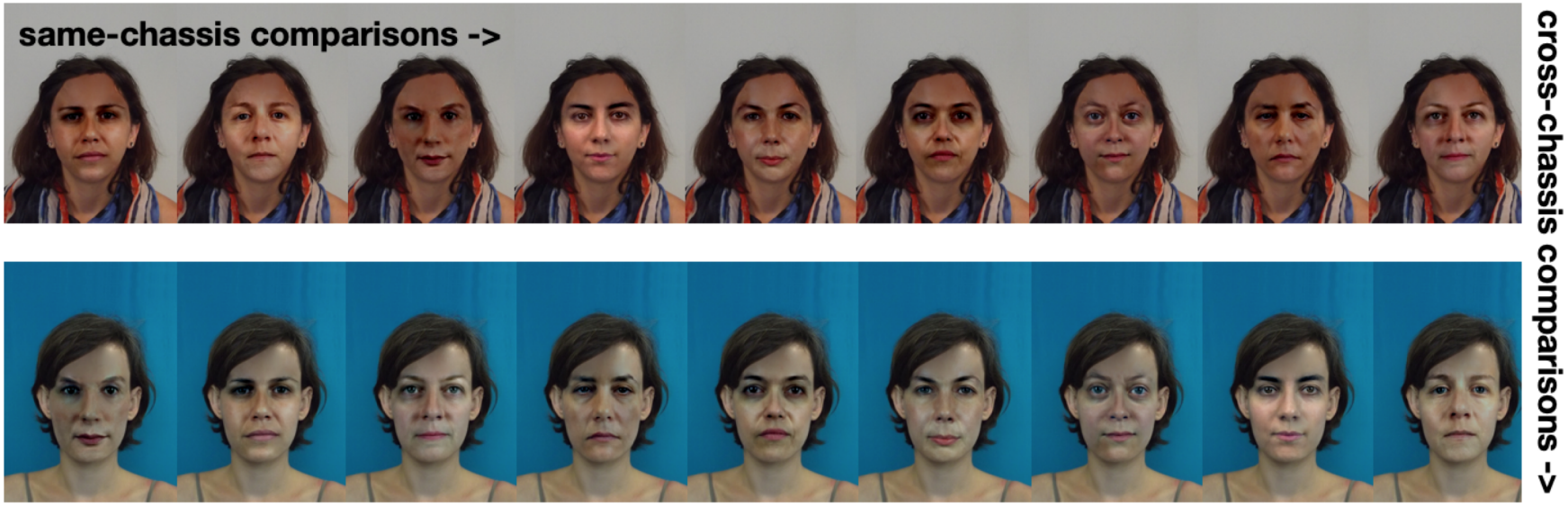


- Such images should emulate comparison scores generated through different chassis and different faces, i.e. true impostor comparisons



Test Dataset

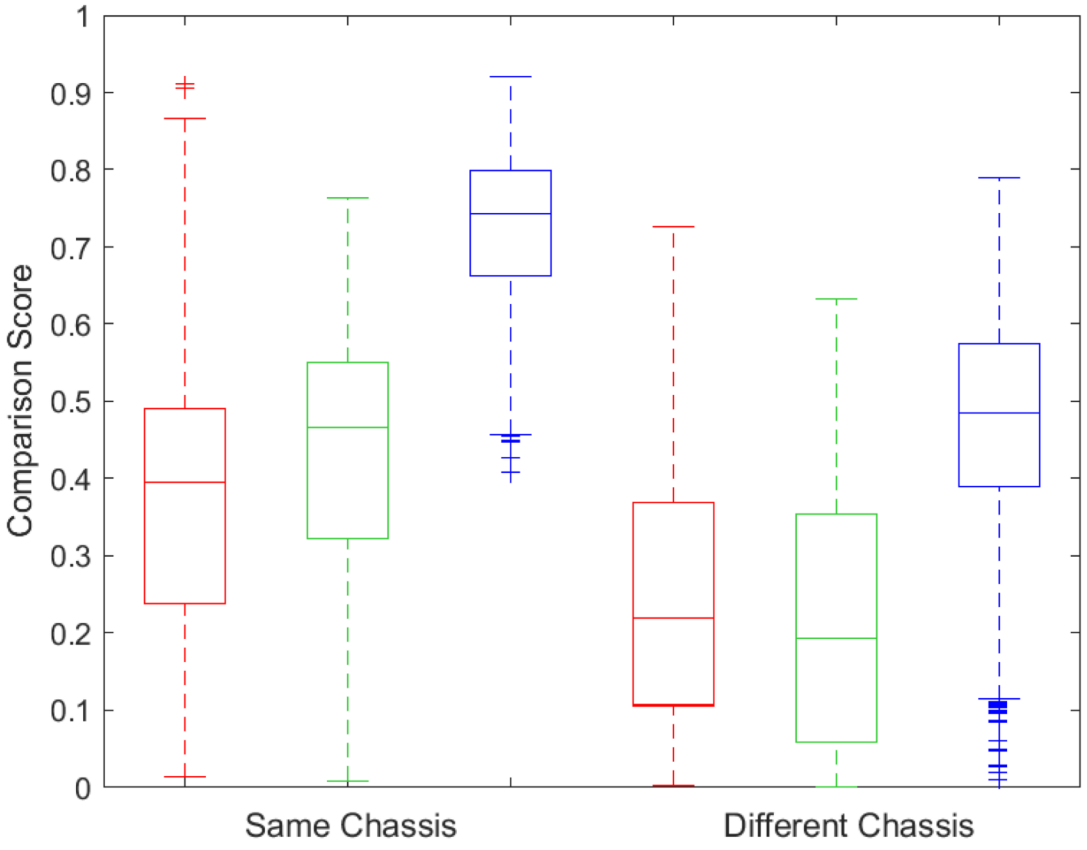
- 10 white female chassis
- 10 synthesized white female faces
- This yielded a total of 100 images
- 4500 comparisons executed
  - 450 same-chassis comparisons, 4050 cross-chassis comparisons
  - A:B comparisons retained, B:A comparisons eliminated



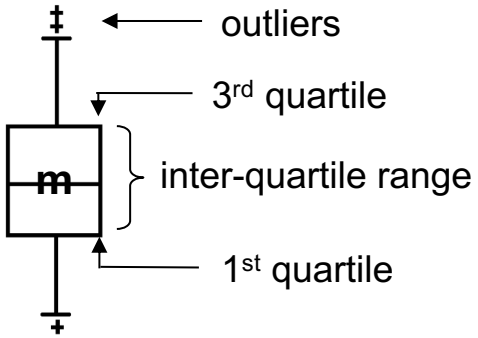
FR Matching Technologies

- Three FR matchers selected for evaluation to gain insights into behavior across different technologies
- Two commercial FR tools (Vendor A and B)
  - Comparison score range: 0-1, higher scores = stronger match
- OpenFace (deep learning, open source)
  - Returns normalized Euclidean distances (d); less distance = stronger match
  - To provide a common scale, complement of the distance (1–d) is reported

Same-Chassis and Different-Chassis Results



-- Vendor A
-- Vendor B
-- OpenFace



Same vs. Different Chassis Analysis

- For all three matchers, the median values are significantly higher for same-chassis comparisons than for different chassis-comparisons
- Variation in distributions between COTS and open source is apparent
  - 1st and 3rd quartiles are about the same for Vendor A and B, whereas OpenFace's are a few tenths higher
  - Vendor A and B have wide inter-quartile ranges relative to OpenFace
  - May be attributable to use of comparison scores vs. Euclidean distance

High-Scoring Same-Chassis Comparisons (1)

Vendor B: 0.7637

High-Scoring Same-Chassis Comparisons (2)

Vendor A: 0.91152

High-Scoring Same-Chassis Comparisons (3)

OpenFace: 0.92072

Same vs. Different Chassis Example

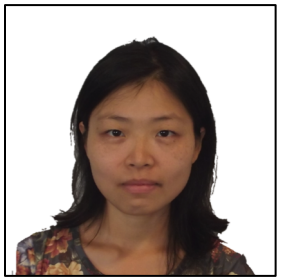
Vendor B: 0.185
OpenFace: 0.597

Scores Increase

Vendor B: 0.449
OpenFace: 0.755

Isolated Chassis Test: Background

- Customer requirements for backgrounds vary
- A further test conducted in which only backgrounds were edited to assess impact on performance

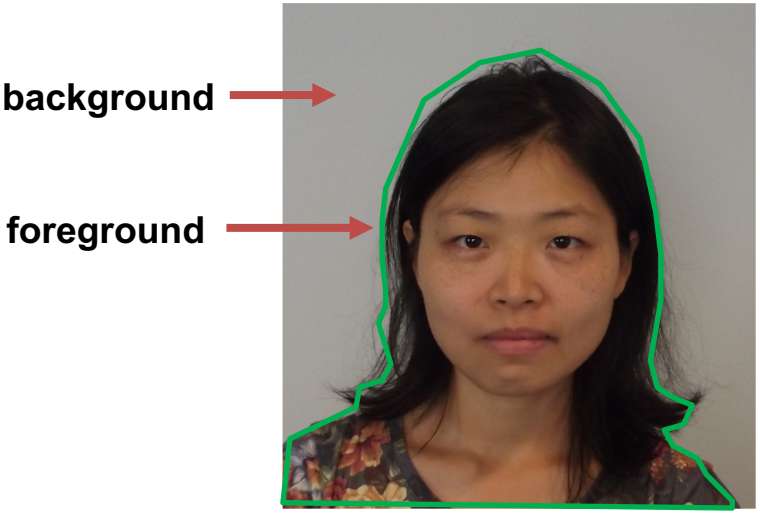

vs.

=> score 1


vs.

=> score 2

Is there an expectation that score 1 > score 2?

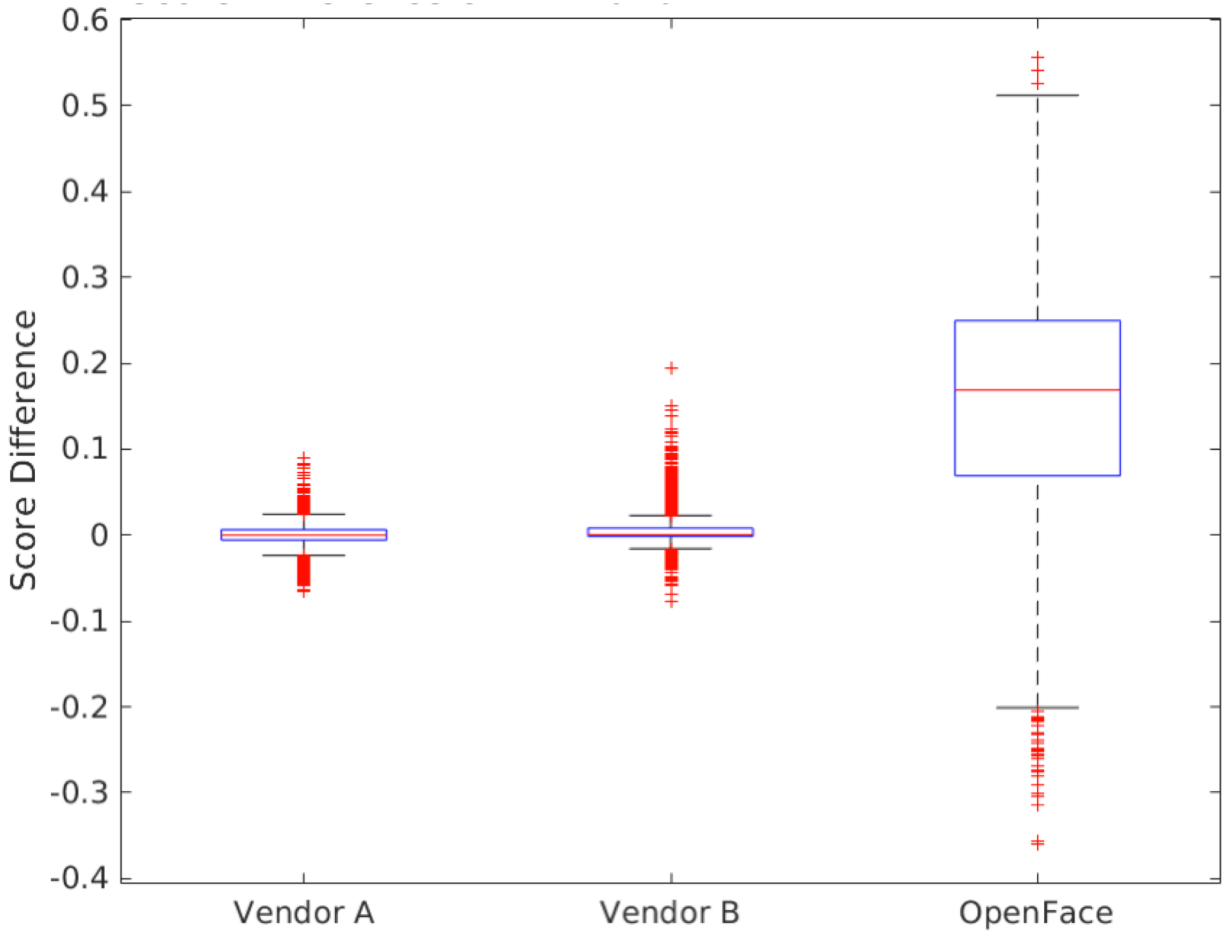


Do comparison pairs with identical color backgrounds generate higher scores than pairs with different color backgrounds?

Isolated Chassis Test: Dataset and Matchers

- 53 real images (faces not modified or synthesized)
- 53 blue background images, 53 white background images
- 5408 impostor comparisons were evaluated
  - 2704 white vs. white comparisons
  - 2704 blue vs. white comparisons
  - A:B comparisons retained, B:A comparisons eliminated
- Same matchers used for this test (2 COTS, OpenFace)

Isolated Chassis Results



$$score\ difference = (white{-}white\ score)\ -(blue{-}white\ score)$$

- Minimal score Δ for COTS vendors
  - Variation mostly within +/- 0.1
  - Median ~0
- OpenFace scores are impacted by background
- Whether this is sufficient to meaningfully impact FMR is TBD

Vendor A Sample Outlier

score difference = 0.0897

Vendor B Sample Outlier

score difference = 0.194

OpenFace Sample Outlier

score difference = 0.557

Conclusions

- Same vs different chassis evaluation results indicate that non-face regions influence FR performance
- Isolated chassis evaluation, however, indicated that the background does not significantly impact FR performance for COTS matchers
- It seems that there are different regions of the image that matchers are considering to varying degrees, such that there is a:
  - Strong concentration in the face region
  - Weak concentration in the background area
  - Hypothesis, medium concentration in the non-face, non-background chassis region

FR Influence Heat Map

Future Work

- Test strong concentration area with real faces
  - There may potentially be something about Novetta's face synthesis that somehow compels the FR system to consider non-face regions
  - Different real faces inserted into the same chassis may have a different outcome
- Validate chassis area as an influential area
  - Develop test to modify the non-face, non-background chassis region such that FR performance can be evaluated
- Improve background editing in the low concentration area
  - In the Isolated Chassis test, all the background values were set to the same RGB pixel intensity
  - Adding pixel intensity variation by emulating typical backdrop photo captures may yield a stronger influence on FR performance